LADDS SAS User's Guide

Table of Contents

Chapter 1	Introduction	2
Chapter 2	Western Blots Assays	3
Chapter 3	Short Form of ELISA, Western Blot, RIBA Blot, etc.	8
Chapter 4	Hepatitis Viruses	10
Chapter 5	HHV-8 Results	11
Chapter 6	HCV Serotype Results	11
Chapter 7	Genotype Database	13
Chapter 8	Backgrnd Table	15
Chapter 9	Samp Table	18
Chapter 10	Subject Table	20
Chapter 11	SQL Statement	21
Chapter 12	Polymerase Chain Reaction (PCR)	22
Chapter 13	SAS Format Libraries	26

1. Introduction

The Laboratory and Demographic Database System (LADDS) contains test results and background information on several hundred thousand subjects in several hundred studies. For the convenience of SAS programmers, a SAS version of the data is kept on line at CIT. This document describes the layouts, locations, and interpretations of these SAS data sets.

MVS FILE NAME	SAS MEMBER NAME	CONTENT
NMG1PGY.FS18.BLOTS	HTLV	HTLV Western Blot Results
	HIV	HIV Western Blot Results
NMG1PGY.FSRSLT.SASDS	HTLV	HTLV Results Summary
	HIV	HIV Results Summary
UAT1PIV.HIV9.HEP.SASDS	BLOT	Hepatitis A, B, and C Blot Results
	SCREEN	Hepatitis A, B, and C Results Summary
NMG1PGY.FSBCK.SASDS	BACKGRND	Background Information
UAT1PIV.HIV9.PCR.SASDS	PCR	PCR Results Summary
UAT1PIV.HIV7.V6FORMAT.SSD		SAS Format Library
UAT1PIV.HIV9.HHV8.SASDS	HHV8	HHV-8 Results Summary
UAT1PIV.HCVSER.SSD	HCVSER	HCV Serotype Results Summary
UAT1PIV.GENO.SSD	GENO	Genotype Results Summary

2. Western Blots Assays

Confirmation fields indicate the overall indication of the presence of the virus. There are 3 confirmation fields in the Blot data set. Each can take one of the following values:

Confirmation Code	Interpretation
+	the specimen tests positive
-	the specimen tests negative
/	indeterminate, or weakly positive
?	The field is relevant, but there is no algorithm to assign a score
(Blank)	The field is not relevant for this virus

The database update software assigns the confirmation scores. For a detailed description on the confirmation algorithms, please refer to "The Scoring of Western and RIBA Blots" section of the LADDS documentation. The table below shows which of the confirmation fields are used, based on the virus(es) the assay is testing.

Virus	VIRUS_ TYP	CONF	CONFI	CONFII	HIV2 Band
HTLV-I	'1'		*		
HTLV-II	'2'			*	
HTLV I/II	'A'	*	*	*	
HIV-1	'3'		*		
HIV-2	'4'			*	
HIV 1/2	'D'		*		*
Hepatitis C	'C'	*			

Blot Assays for Human T-Cell Lymphotropic Viruses (HTLV)

Filename: NMG1PGY.FS18.BLOTS
SAS Internal name: HTLV
Contents: HTLV Western Blots Results

Variable	Type/Len	Description
BEST	Char 1	Best (Set to "*" for the most recent test date of this sample id, virus type, result type, and lab)
CONF	Char 1	HTLV Confirmation
CONFI	Char 1	Type I Confirmation
CONFII	Char 1	Type II Confirmation
DT_DRAWN	Char 6	Draw date
LAB	Char 2	Testing Lab
P15	Char 1	P-15 Score
P19	Char 1	P-19 Score
P21	Char 1	P-21 Score
P22	Char 1	P-22 Score
P24	Char 1	P-24 Score
P26	Char 1	P-26 Score
P28	Char 1	P-28 Score
P32	Char 1	P-32 Score
P36	Char 1	P-36 Score
P38	Char 1	P-38 Score
P40	Char 1	P-40 Score
P42	Char 1	P-42 Score
P46	Char 1	P-46 Score
P53	Char 1	P-53 Score
P55	Char 1	P-55 Score
P21E	Char 1	P21E Score
RACE	Char 1	Race
RCVD_DT	Char 6	Date data received in LADDS
RGP461	Char 1	RGP-46I Score
RGP462	Char 1	RGP-46II Score

RSLT_TYP	Char 3	Assay Type
SAMPL_ID	Char 9	Sample Id
SC	Char 1	Serum Control Indicator Score
SEX	Char 1	Sex
STUDY_ID	Char 3	Study Id
SUBJ_ID	Char11	Subject Id
TESTDTS	Num 4	Test Date
TRAY	Char 4	Test Tray number
VIAL_NO	Char 4	Specimen Vial number
VIRUSTYP	Char 1	Virus type

Blot Assays for Human Immunodeficiency Viruses (HIV)

Filename: NMG1PGY.FS18.BLOTS
SAS Internal name: HIV
Contents: HIV Western Blots Results

Variable	Type/Len	Description
BEST	Char 1	Best (Set to "*" for the most recent test date of this sample id, virus type, result type, and lab)
CONF	Char 1	Overall Confirmation
CONFI	Char 1	Type I Confirmation
CONFII	Char 1	Type II Confirmation
DT_DRAWN	Char 6	Draw date
ENVOLIG	Char 1	Env Oligomers
GP34	Char 1	GP-34 Score
GP41	Char 1	GP-41 Score
GP105	Char 1	GP-105 Score
GP120	Char 1	GP-120 Score
GP140	Char 1	GP-140 Score
GP160	Char 1	GP-160 Score
GP34PRIM	Char 1	GP-34 Primer
HIV2	Char 1	HIV-2 Indicator Score
LAB	Char 2	Test Lab
P15	Char 1	P-15 Score

P16	Char 1	P-16 Score
P17	Char 1	P-17 Score
P24	Char 1	P-24 Score
P26	Char 1	P-26 Score
P31	Char 1	P-31 Score
P36	Char 1	P-36 Score
P39	Char 1	P-39 Score
P45	Char 1	P-45 Score
P51	Char 1	P-51 Score
P55	Char 1	P-55 Score
P56	Char 1	P-56 Score
P58	Char 1	P-58 Score
P66	Char 1	P-66 Score
P68	Char 1	P-68 Score
RACE	Char 1	Race
RCVD_DT	Char 6	Date data received in LADDS
RSLT_TYP	Char 3	Assay type
SAMPL_ID	Char 8	Sample Id
SC	Char 1	Serum Control Indicator Score
SEX	Char 1	Sex
STUDY_ID	Char 3	Study Id
SUBJ_ID	Char11	Subject Id
TESTDTS	Num 4	(SAS) Test date
TRAY	Char 4	Test Tray number
VIAL_NO	Char 4	Specimen Vial number
VIRUSTYP	Char 1	Virus Type

Blot Assays for Hepatitis Viruses

Filename: UAT1PIV.HIV9.HEP.SASDS
SAS Internal name: BLOT
Contents: Western Blot Results for Hepatitis A, B and C

Variable	Type/Len	Description
BEST	Char 1	Best (Set to "*" for the most recent test date of this sample id, virus type, result type, and lab)
C22	Char 1	C-22 Score
C33	Char 1	C-33 Score
C100	Char 1	C-100 Score
C511	Char 1	C-511 Score
CONF	Char 1	Overall Confirmation
DT_DRAWN	Char 6	Draw date
LAB	Char 2	Lab
NS5	Char 1	NS5 Score
RACE	Char 1	Race
RCVD_DT	Char 6	Date data received in LADDS
RSLT_TYP	Char 3	Assay type
SAMPL_ID	Char 9	Sample Id
SEX	Char 1	Sex
SOD	Char 1	Super Oxide Dismutase
STUDY_ID	Char 3	Study Id
SUBJ_ID	Char 11	Subject Id
TESTDTS	Num 4	Test date
TRAY	Char 4	Test Tray number
VIAL_NO	Char 4	Specimen Vial number
VIRUSTYP	Char 1	Virus Type

3. Short Form of ELISA, Western Blots

Human T Lymphotropic Viruses (HTLV) Type '1', '2', and 'A'

Filename: NMG1PGY.FSRSLT.SASDS
SAS Internal name: HTLV
Contents: HTLV Results Summary

Variable	Type/Len	Description
BEST	Char 1	Best (Set to "*" for the most recent test date of this sample id, virus type, result type, and lab)
CONF	Char 1	HTLV Confirmation
CONFI	Char 1	Type I Confirmation
CONFII	Char 1	Type II Confirmation
DRAW_DT	Date	Draw Date
LAB	Char 2	Testing Lab
RACE	Char 1	Race
RATIO	Num 4	1ST numeric result
REAL2	Num 4	2ND numeric result
REAL3	Num 4	3RD numeric result
REAL4	Num 4	4TH numeric result
RSLT_TYP	Char 3	Assay type
SAMPL_ID	Char 9	Sample Id
SEX	Char 1	Sex
SOURCE	Char 1	Source file (Blot or Result file)
STUDY_ID	Char 3	Study Id
SUBJ_ID	Char 11	Subject Id
TEST_DT	Date	Test date
VIAL_NO	Char 4	Specimen Vial number
VIRUSTYP	Char 1	Virus type

Human Immunodeficiency Viruses (HIV) Type '3', '4', and 'D'

Filename: NMG1PGY.FSRSLT.SASDS
SAS Internal name: HIV
Contents: HIV Results Summary

Variable	Type/Len	Description
BEST	Char 1	Best (Set to "*" for the most recent test date of this sample id, virus type, result type, and lab)
CONF	Char 1	Overall Confirmation
CONFI	Char 1	Type I Confirmation
CONFII	Char 1	Type II Confirmation
DRAW_DT	Date	Draw Date
HIV2	Char 1	HIV2 Indicator Score
LAB	Char 2	Testing Lab
MOD	Char 1	Misc. modifier
RACE	Char 1	Race
RATIO	Num 4	1ST numeric result
REAL2	Num 4	2ND numeric result
REAL3	Num 4	3RD numeric result
REAL4	Num 4	4TH numeric result
RSLT_TYP	Char 3	Assay Type
SAMPL_ID	Char 9	Sample Id
SEX	Char 1	Sex
SOURCE	Char 1	Source file (Blot or Result file)
STUDY_ID	Char 3	Study Id
SUBJ_ID	Char 11	Subject Id
TEST_DT	Date	Test Date
VIAL_NO	Char 4	Specimen Vial Number
VIRUSTYP	Char 1	Virus Type

4. Hepatitis Viruses

Filename: UAT1PIV.HIV9.HEP.SASDS
SAS Internal name: SCREEN
Contents: Results Summary for Hepatitis A, B and C viruses

Variable	Type/Len	Description
BEST	Char 1	Best (Set to "*" for the most recent test date of this sample id, virus type, result type, and lab)
CONF	Char 1	1st POS/NEG Score
DT_DRAWN	Char 6	Draw date (yymmdd)
LAB	Char 2	Testing Lab
RACE	Char 1	Race
RATIO	Num 4	1st Numeric result
RCVD_DT	Char 6	Date data received in LADDS
REAL2	Num 4	2nd Numeric result
REAL3	Num 4	3rd Numeric result
REAL4	Num 4	4th Numeric result
RSLT_TYP	Char 3	Assay type
SAMPL_ID	Char 9	Sample Id
SEX	Char 1	Sex
SOURCE	Char 1	Source file (Blot or Result file)
STUDY_ID	Char 3	Study Id
SUBJ_ID	Char 11	Subject Id
TESTDTS	Num 4	Test date
VIAL_NO	Char 4	Specimen Vial number
VIRUSTYP	Char 1	Virus type

5. HHV-8 Results

Filename: UAT1PIV.HIV9.HHV8.SASDS
SAS Internal name: HHV8
Contents: Results Summary for Human Herpes Virus-8

Variable	Type/Len	Description
BEST	Char 1	BEST RESULT
CONF	Char 1	OVERALL CONFIRMATION
DT_DRAWN	Char 6	DATE DRAWN (yymmdd)
LAB	Char 2	TESTING LAB
RACE	Char 1	RACE
RATIO	Num 4	1st NUMERIC RESULT
RCVD_DT	Char 6	DATE IN DATABASE
REAL2	Num 4	2nd NUMERIC RESULT
RSLT_TYP	Char 3	ASSAY TYPE CODE
SAMPL_ID	Char 9	SAMPL_ID
SEX	Char 1	SEX
SOURCE	Char 1	SOURCE
STUDY_ID	Char 3	STUDY ID
SUBJ_ID	Char 11	SUBJECT ID
TESTDTS	Num 4	TEST DATE
VIAL_NO	Char 4	VIAL NUMBER
VIRUSTYP	Char 1	VIRUS TYPE

6. HCV Serotype Results

This SAS data set contains test results from a single assay used to discriminate between HCV infections with serotypes 1,2 or 3 to aid in the diagnosis and treatment of patients with hepatitis C.

The CHIRON*RIBA*HCV Serotyping Strip Immunoblot Assay (SIA) is an in vitro qualitative Enzyme Immuno-Assay (EIA) to detect serotype-specific antibodies present in human serum or plasma. It is intended as a supplemental assay to be used to test speciments found to be reactive using an anti-HCV screening procedure and subsequently

confirmed as positive by CHIRON*RIBA*HCV 2.0 SIA or CHIRON*RIBA*HCV 3.0 SIA.

The CHIRON*RIBA*HCV Serotyping SIA uses an immunoblot EIA technique for the detection of antibodies to the serotype-specific epitopes of the hepatitis C virus. This assay utilizes 8 synthetic HCV-encoded peptides. All of these are highly serotype-specific, with the exception of the serotype 1 core peptide which shows significant cross-reactivity with serotype 3 seciments. Five peptides are from the serotype-specific sequences of the NS-4 nonstructural coding region of the genome, and three peptides are from the core region of the different HCV isolates.

The peptides are coated individually or in combinations of 2 peptides in each band of the serotyping strip. Band 1 contains the serotype specific type 1a and 1b epitopes, band 2 contains the type 2a and 2b epitopes, and band 3 contains the type 3 epitope from the NS4 region of the HCV genome. Band 4 contains the type1 epitope, while band 5 contains type 2 epitopes of the Japanese and Italian HCV variant from the core region.

Filename: UAT1PIV.HCVSER.SSD
SAS Internal name: HCVSER
Contents: Results Summary for HCV Serotype assays

Variable Name	Data Type	Length	Description
STUDY_ID	char	3	A unique identifier for each study. Must exist in study table.
SUBJ_ID	char	11	A unique identifier for each subject. Must exist in subject table.
SAMPL_ID	char	9	The repository assigns a unique sample identifier is assigned to each specimen collected from a subject. Mandatory. See SAMPL_ID in SAMP Table.
VIAL_NO	char	3	Specimen Vial Sequence number. Either three digits or blank.
LAB	char	2	Always 'P' SAIC
VIRUSTYP	char	1	LADDS virus code. Always 'C'
RSLT_TYP	char	4	LADDS assay code. Always 'RCS'
TEST_DT	num	4	The date sample was tested (SAS date)
TRAY	char	4	Specimen Tray number - a laboratory code.
QUALITY	char	1	Quality of the data. Ignore this, always Blank.
EDITTYPE	char	1	Type of Band scoring. Always 'A' meaning that all band scores are on an eleven point scale from 0(weakest) to 10(strongest) used at the SAIC laboratory. In the file

			these scores are stored as '0'-'9' and 'A' for 10.
DILUTION	num	5	The dilution of the sample. Values are '10', '33', '132', and '660'.
CNTL1	char	1	Low Level I Human IGG control band. Internal control.
CNTL2	char	1	High Level II Human IGG control band. Internal control.
NS41	char	1	Reactivity to NS4 serotype specific type 1a and 1b epitopes.
NS42	char	1	Reactivity to NS4 serotype specific type 2a and 2b epitopes.
NS43	char	1	Reactivity to NS4 type 3 epitope.
CORE1	char	1	Reactivity to CORE epitope 1.
CORE2	char	1	Reactivity to epitopes of the Japanese and Italian HVC variant CORE regions.
DT_DRAWN	char	6	The date (yymmdd) the sample was draw.
SEROTYP	char	3	Serotyping Result: • '1' = Serotype 1 • '2' = Serotype 2 • '3' = Serotype 3 • '13' = Serotype 1 or 3 • 'UTY' = Untyped • 'NEG' = Negative (applies to Negative Control only) • 'POS' = Positive (applies to Positive Control only) • 'DIL' = Dilute sample and re-test.
T_STAMP	char	8	The date (yyyymmdd) this record entered the database.

7. Genotype Data Base

Chemokines form a large family of small proteins with four conserved cysteines lined to disulfide bonds. They are divided into subfamilies based on the motif formed by the first two cysteines. Thus, for the CXC chemokines, i.e. interleukin-8 (IL-8), the first two cysteines are separated by one amino acids, and the for CC chemokines, for example, CCR5, the first two cysteines are adjacent. Three variations on this are a CXXXC (CX3C) chemokine fractalkine, a C chemokine lymphotactin, which has only one cysteine pair, and several CC chemokines that have a third disulfide bridge, i.e. SLC also known as C6kine. Most chemokines belong to either the CC or CXC category. The backbone fold of the chemokine is the key determinant of its function, particularly the N-terminal end.

Chemokines act via heptahelical receptors coupled to the GTP-binding proteins. They induce chemotaxis and several transient responses such as intracellular Ca2+ changes, the production of oxygen radicals and bioactive lipids, and the release of storage gradule contents. Many chemokines are produced by tissue cells and infiltrating leukocytes under pathological conditions, inflammation in particular. T-lymphocytes express the largest variety of chemokine receptors and respond to an impressive number of different chemokines. The polymorphic expression (either heterozygous or homozygous alleles-referred to as GENOTYPE in the dataset) of certain chemokines such as CCR5-delta 32 and CCR2-I64 have been shown to be important in the invasion of HIV into T-lymphocytes, and hence to AIDS progression.

Abstracted from MEDLINE and http://www.aegis.com

The SAS data set is UAT1PIV.GENO.SSD member GENO. **In general, there is one genotype result in the database (in the case of retesting, it will be the latest result) per subject per locus, however when there are different labs testing for the same locus, the genotype result from each laboratory will be stored.**

Variable Name	Data Type	Length	Description
GENOTYPE	char	3	1|1="wild type" normal on both alleles 1|2=a wild type allele and mutant allele 2|2=two mutant alleles
HGAL	char	7	Laboratory for Genetic Diversity (LGD-Dr. Steve O'Brien) laboratory identification number.
LAB	char	2	The testing laboratory, ie, MC=Mary Carrington (LGD) and P=SAIC. A missing value denotes information is unavailable.
LOCUS	char	3	Gene Locus
SAMPL_ID	char	9	The repository assigns a unique sample identifier to each specimen collected from a subject. MANDATORY.
STUDY_ID	char	3	A unique identifier for each study.
SUBJ_ID	char	11	An identifier for a subject.
TEST_DT	date	8	The sample test date (YYYYMMDD format on transaction file and stored as a SAS date in the data base).
LADDS_DT	date	8	The genotype database update date (YYYYMMDD format on transaction file and stored as a SAS date in the database).

PP_SET	char	3	Primer probe set allele detection code used by the SAIC lab (LAB=P) only. This variable is missing for all other labs. New codes will be created as the detection technology improves.

8. Backgrnd Table

A DB2 table of specimen information - UAT1PIV.BACKGRND

A DB2 view of the Subject and Sample tables - UAT1PIV.BACKGRND		
BIRTH_DT	Date	Birth Date. (See BIRTH_DF)
BIRTH_DF	Short Integer	The Birth Date Flag indicates if the date was imputed. This flag and the imputations are not collected in the field - They are maintained by the database software. • Complete dates have this flag set to 0. • Dates that were just missing only the day have this flag set to 1 and the day set to 15. • Dates that were missing month and day have this flag set to 2 and the month and day set to 0702. • Dates that were totally missing have this flag set to 3 and the date set to NULL
BIRTH_PL	Char (8)	Birth Place
CRESAMP	Date	The data that this sample was added to the data base This date is not collected in the field - it is maintained by the database software.
CRESUBJ	Date	The date that this subject was added the data base. This date is not collected in the field - it is maintained by the database software.
DIAG1	Char (10)	Primary diagnosis code If ICD8 or ICD9 remove the decimal point. If ICD-O 1 or 2 or 3 Topography in cols 1/4 (remove decimal points) If ICD-O 1 or 2 or 3 Histology in cols 5/8 Topology. If ICD-O 1 or 2 or 3 Behavior in cols 9. If ICD-O 1 or 2 or 3 Grade or differentiation in cols 10. Or special coding for Lymphomas. See ICD-O documentation.
DIAG1_CD	Char (1)	The coding scheme used for DIAG1

		• 8 = ICD8 • 9 = ICD9 • 1 = ICDO version 1 • 2 = ICDO version 2 • 3 = ICDO version 3 • L = Paul Levine code
DIAG1_DT	Date	The date of DIAG1 (Same rules as BIRTH_DT)
DIAG1_DF	Num (8)	DIAG1 Date Flag (Same rules as BIRTH_DF)
DIAG2	Char (10)	Secondary Diagnosis code
DIAG2_CD	Char (1)	The coding scheme used for DIAG2 • 8 = ICD8 • 9 = ICD9 • 1 = ICDO version 1 • 2 = ICDO version 2 • 3 = ICDO version 3 • L = Paul Levine code
DIAG2_DT	Date	The DIAG2 Date (Same rules as BIRTH_DT)
DIAG2_DF	Num (8)	DAIG2 Date Flag (Same rules as BIRTH_DF)
DRAW_DT	Date	Date the specimen was collected. (Same rules as BIRTH_DT)
DRAW_DF	Num (8)	Draw Date Flag.(Same rules as BIRTH_DF)
DRAW_HR	Char (2)	Draw Hour ("00" = midnight, "01" = 1 a.m. to "23" = 11 p.m.)
ETHN1	Char (2)	The Subject's 1st Ethnicity
ETHN2	Char (2)	The Subject's 2nd Ethnicity
ETHN3	Char (2)	The Subject's 3rd Ethnicity
ETHN4	Char (2)	The Subject's 4th Ethnicity
MODSAMP	Date	The date that sample linked information was last modified for this sample
MODSUBJ	Date	The date that subject linked information was last modified for this subject
NBR_PREG	Char (2)	Number of Pregnancies
RACE	Char (1)	The subject's Race

RESIDNCE	Char (8)	Current Residence
SAMPL_ID	Char (9)	Sample Id
SEX	Char (1)	Sex Code • blank = unknown • 1 = male • 2 = female • 9 = not disclosed
STUDY_ID	Char (4)	Study Id
SUBJ_ID	Char (11)	Subject Id left justified letters, numbers, or special characters "_-*#+=.@&!<>"
TRANF_DT	Date	Date of Last Blood Transfusion. (Same rules as BIRTH_DT)
TRANF_DF	Num (8)	Last Blood Transfusion Date Flag (Same rules as BIRTH_DF)
YRS_RES	Char (2)	Years at the current residence
ZIPCODE	Char (5)	Subject's Zip Code

9. Samp Table

A DB2 table of specimen information - UAT1PIV.SAMP

Variable	Type	Description
DIAG1	Char (10)	Primary diagnosis code If ICD8 or ICD9 remove the decimal point. If ICD-O 1 or 2 or 3 Topography in cols 1/4 (remove decimal points) If ICD-O 1 or 2 or 3 Histology in cols 5/8 Topology. If ICD-O 1 or 2 or 3 Behavior in cols 9. If ICD-O 1 or 2 or 3 Grade or differentiation in cols 10. Or special coding for Lymphomas. See ICD-O documentation.
DIAG1_CD	Char (1)	The coding scheme used for DIAG1 • 8 = ICD8 • 9 = ICD9 • 1 = ICDO version 1 • 2 = ICDO version 2 • 3 = ICDO version 3 • L = Paul Levine code
DIAG1_DT	Date	1st Diagnosis Date (See DIAG1_DF)
DIAG1_DF	Small int	1st Diagnosis Date Flag: Codes if original date was complete. Complete dates have this flag set to 0. Dates that were just missing the day have this flag set to 1 and the day set to 15. Dates that were missing month and day have this flag set to 2 and the month and day set to 0702. Dates that were totally missing have this flag set to 3 and the date set to NULL
DIAG2	Char (10)	Secondary diagnosis code
DIAG2_CD	Char (1)	The coding scheme used for DIAG2 • 8 = ICD8 • 9 = ICD9 • 1 = ICDO version 1 • 2 = ICDO version 2 • 3 = ICDO version 3 • L = Paul Levine code
DIAG2_DT	Date	2nd Diagnosis Date (Same rules as DIAG1_DT)
DIAG2_DF	Smallint	2nd Diagnosis Flag (Same rules as DIAG1_DF)
DRAW_DT	Date	Date the specimen was collected. (Same rules as DIAG1_DT)

DRAW_DF	Smallint	Draw Date Flag: (Same rules as DIAG1_DF)
DRAW_HR	Char (2)	Draw Hour
FIRST_DT	Date	Date Sample in LADDS
MOD_DT	Date	Date Sample Modified in LADDS
NBR_PREG	Char (2)	Number of Pregnancies
RESIDNCE	Char (8)	Residence
SAMPL_ID	Char (9)	Sample Id
STUDY_ID	Char (4)	Study Id
SUBJ_ID	Char (11)	Subject Id
TRANF_DT	Date	Date of Last Blood Transfusion. (Same rules as DIAG1_DT)
TRANF_DF	Small int	Last Blood Transfusion Date Flag (Same rules as DIAG1_DF)
YRS_RES	Char (2)	Year at Residence
ZIPCODE	Char (5)	Zipcode of Residence

10. Subject Table

A DB2 view of the SUBJ table - UAT1PIV.SUBJECT

Variable	Type	Description
BIRTH_DT	Date	Birth Date. (See BIRTH_DF)
BIRTH_DF	Small int	Birth Date Flag: Codes if original date was complete. Complete dates have this flag set to 0. Dates that were just missing the day have this flag set to 1 and the day set to 15. Dates that were missing month and day have this flag set to 2 and the month and day set to 0702. Dates that were totally missing have this flag set to 3 and the date set to NULL
BIRTH_PL	Char (8)	Birth Place
ETHN1	Char (2)	1st Ethnicity
ETHN2	Char (2)	2nd Ethnicity
ETHN3	Char (2)	3rd Ethnicity
ETHN4	Char (2)	4th Ethnicity
F_NAME	Char (8)	First Name
FIRST_DT	Date	Date Subject Entered LADDS
L_NAME	Char (12)	Last Name
M_NAME	Char (8)	Middle Name
MOD_DT	Date	Date Subject Changed in LADDS
RACE	Char (1)	Race code
SEX	Char (1)	Sex code
STUDY_ID	Char (4)	Study Id
SUBJ_ID	Char (11)	Subject Id

11. SQL Statement

Example of using SQL to access the view UAT1PIV.BACKGRND to create a SAS data set named BACKGRND containing SIB subjects whose ids start with "M14".

```
OPTIONS NOCENTER LS=80 OBS=MAX SYMBOLGEN ;

PROC SQL ;
CONNECT TO DB2 (SSID=DSNP) ;
CREATE TABLE BACKGRND AS
SELECT * FROM CONNECTION TO DB2
(
SELECT STUDY_ID, SUBJ_ID, SAMPL_ID, DRAW_DT,
RACE, SEX
FROM UAT1PIV.BACKGRND
WHERE STUDY_ID = 'SIB' AND
SUBJ_ID LIKE 'M14%'
ORDER BY STUDY_ID, SUBJ_ID
);
%PUT &SQLXMSG ;
%PUT &SQLXRC ;
DISCONNECT FROM DB2 ;
```

12. Polymerase Chain Reaction (PCR)

In 1985 Kary Mullis developed the PCR technique for which he was awarded the 1993 Nobel Prize in chemistry. PCR is a laboratory procedure in which a particular DNA segment (sequence) present in a heterogeneous population of DNA fragments, is replicated many times (amplified) to produce a large number of exact DNA copies. The polymerase chain reaction (PCR) is a valuable technique that is widely used in a number of established laboratory/clinical applications but also shows promise for newly emerging applications. These include DNA fingerprinting, bacterial and viral sequence detection, molecular cloning, gene therapy, and diagnosis of genetic disorders.

To amplify a segment of DNA using PCR, the sample is first heated so the DNA denatures, or separates into two pieces of single-stranded DNA. Next, an enzyme called Taq polymerase synthesizes, or makes, two new strands of DNA, using the original strands as templates; this process results in the duplication of the DNA, with each of the new molecules containing one old and one new strand of DNA. The cycle of denaturing and synthesizing new DNA is repeated as many as 30 or 40 times, leading to more than 1 billion exact copies of the original segment of DNA. To amplify RNA rather than DNA sequences, a DNA copy (called cDNA) of the RNA can be made by reverse transcription prior to PCR.

Each PCR cycle requires rapid changes in temperature to allow for denaturation (typically at 93-95 degrees C), annealing (usually 50-55 degrees C) and subsequent DNA synthesis (typically at 60-65 degrees C) to take place. The exact temperatures selected for each of the steps varies depending upon the sequences of the target DNA/RNA and primers, and the reaction is carried out in a tube containing a small volume of liquid, typically 50 microliters. A thermocycler is a programmable laboratory instrument that automates the entire PCR temperature cycling process so it can be completed in a few hours. The thermocycler contains a heat block capable of maintaining a precise temperature and rapidly changing temperature after a few minutes or even a few seconds.

Qualitative PCR provides a '+' (positive), '-' (negative) or '/' (indeterminate) answer (variable=CONF) as to whether there is a detectable level of DNA/RNA in a sample, but does not provide information about the amount of DNA/RNA present. To accurately quantify the amount of target DNA/RNA present in a sample, an approach called Real-Time (quantitative) PCR is used. In principal this technology is the same as conventional PCR, except that the amount of DNA/RNA amplified during each cycle of the PCR process is monitored using fluorescent dyes. This requires a specialized instrument (such as an ABI PRISM 7700 or 7900) capable of detecting various fluorescent dyes and measuring them at very short intervals. This monitoring yields an amplification curve for each sample and by comparing the sample to a known standard, precise information on the quantity of target DNA/RNA in the sample can be determined.

PCR Data Collection Specifications

Quantitative PCR (ASSAY PATTERN= "PCRN") results are collected in 3 formats:

1) Copies per ml
2) Copies detected (TAQMAN) - ERV-3 assay
3) Copies per 100 Lymphocytes (%) (TAQMAN) - Beta-actin assay

The results are distinguished in the database as follows:

1) If the ASSAY TYPE is "PCR" the results represent copies per ml.
2) If the ASSAY TYPE is “PCR 7700" the results represent copies detected
3) If the ASSAY TYPE is "PCR 7700A" the results represent copies per 100 lymphocytes

In all cases the following data are required:

1) Copies (as above)
2) Cutoff (the smallest number of copies the assay can reliably detect)
3) Confirmation ("+" if the copies are equal to or greater than the cutoff else "-") (If the testing lab is SAIC and it is a Quantitative TAQMAN, the RNA_PCR may be –1 when the conf = “+” and the copies are less than the cutoff. This scenario is ‘a semiquanitative positive result’).

Sections abstracted from the NIH Office of Science Education and Research website: http://www.nhgri.nih.gov. Editorial contributions also provided by Betty Conde from the NCI Laboratory in Frederick, MD.

FILE NAME: UAT1PIV.HIV9.PCR.SASDS
SAS DATASET NAME: PCR
CONTAINS: PCR Results

Variable	Type/Len	Description
AVG_OD	Num 8	The average OD of the sample
BSIBATCH	Char 12	The BSI sample requisition number
CONF	Char 1	Was DNA or RNA detected?
CUTOFF	Num 8	The least number of copies/ml the assay can reliably detect.
DILUTION	Num 8	The assay dilution factor (null=not diluted,10=10:1,etc.)
DRAW_DF	Num 8	The Draw Date Flag indicates if the date was imputed. This flag and the imputations are not collected in the field - They are maintained by the database software.

		• Complete dates have this flag set to 0. • Dates that were just missing only the day have this flag set to 1 and the day set to 15. • Dates that were missing month and day have this flag set to 2 and the month and day set to 0702. • Dates that were totally missing have this flag set to 3 and the date set to NULL
DRAW_DT	Date	The date sample was drawn.
EXP_ID	Char 8	SAIC Experiment ID
IUNITS	Num 8	International units of RNA or DNA detected
LAB	Char 2	Laboratory where testing was performed.
LADDS_DT	Date	The database transaction date
MAT_CODE	Char 3	Material Code
OD1	Num 8	The optical density of well 1
OD2	Num 8	The optical density of well 2
OD260	Num 8	OD260 wavelength
OD280	Num 8	OD280 wavelength
PCRSCORE	Char 1	The SAIC confirmation score
PP_SET	Char 3	SAIC Primer Probe Code
PX1	Num 8	The first copy count
PX2	Num 8	The second copy count
PX3	Num 8	The third copy count
RACE	Char 1	LADDS Race code
RATIO	Num 8	Average OD of sample/average OD of negative controls
RNA_PCR	Num 8	The copies/ml of DNA or RNA detected
RSLT_TYP	Char 3	The Assay Code
SAMPL_ID	Char 9	The sample ID number
SEX	Char 1	The gender of the subject
STUDY_ID	Char 4	LADDS study identification code
SUBJ_ID	Char 11	LADDS subject identification number
TEST_DT	Date	The date the sample was tested.
VIAL_NO	Char 4	The vial number of the sample
VIRUSTYP	Char 1	The virus code
VOLUME	Num 8	The volume in ul

YIELD	Num 8	The yield in ug

13. SAS Format Libraries

FILE NAME: UAT1PIV.HIV7.V6FORMAT.SSD - formats ending with '' are available by choosing option "O", "2" of the VEB Utilities.*

Format Name	Formats	Example of Fields that would be Relevant for this Format</th
$BANDFMT*	Western Blot Band Scores	P17 P24
$BMATFMT	BSI Material Types	MAT_TYPE
$BMODFMT	BSI Modifier Codes	MODIFIER_LIST
$BVIALFMT	BSI Vial Status Codes	VIAL_STATUS
$BVOLFMT	BSI Volume Unit Codes	VOLUME
$CONFMT*	Confirmations	CONF CONFI CONFII CONF1 CONF2
$ETHN_B	Ethnic Groups	ETHN1 ETHN2
$FAM	"Family" Studies	STUDY_ID
$GENDER*	Sex (Male/Female/Unknown)	SEX
$HEMFMT	Hemophilia Studies	STUDY_ID
$LABF	Laboratory Code and Abbreviation	LAB
$LABFMT*	Laboratory Code and Name	LAB
$LICSTES	Assay test license ('Yes'/'No')	LICENSE
$MATF	PCR Material Code	MAT_CODE
$PCRFMT	PCR confirmations	PCRSCORE
$PPSET	Primer/Probe Id	PP_SET
$QUALTYF	Quality of Data	QUALITY
$RACE_F*	Race	RACE
$RLSTVAL	Screen Assay Conf Pattern	PAT2 PAT9
$STUDYFMT*	Study Id	STUDY_ID
$TT2FMT*	Assay Code and Name	RSLT_TYP
$TYFMT	Assay Code and Abbreviation	RSLT_TYP
$VIRUSF	Virus Type Code	VIRUS
$WHOSTUD	Study Investigator	STUDY_ID

FILE NAME: UAT1PIV.HIV7.GEOCODE.SASFMT

Format Name	Formats	Example of Fields that would be Relevant for this Format
$CTYCTRY	City and country	BIRTH_PL, RESIDNCE

FILE NAME: UAT1PIV.HIV7.VEB.ICD9CM.V6FORMAT.SSD

Format Name	Formats	Example of Fields that would be Relevant for this Format
$ICD9CM	ICD9-CM (excluding procedures) codes	DIAG1, DIAG2

FILE NAME: UAT1PIV.HIV7.VEB.ICD.V6FORMAT.SSD

Format Name	Formats	Example of Fields that would be Relevant for this Format
$ICD9O	Combination of ICD-9 and ICD-O codes: topology	DIAG1, DIAG2
$MORPH	Combination of ICD-9 and ICD-O codes: morphology	DIAG1, DIAG2
$DIGFMT	Modified ICD-O 6th digit code for grade differentiation/cell designation/position.	DIAG1, DIAG2